

# 水文监测中大数据技术的软件应用分析

王 爽

(河南省南阳水文水资源勘测局, 河南 南阳 473000)

**摘要:** 时代技术在进步, 带动了水利信息建设以较快的速度发展, 形成了大规模的水利数据, 将大数据技术合理引进, 促进了水利数据处理的效果。本文主要围绕着大数据技术来展开, 在水文监测处理中, 研究水利数据的特点, 探究数据分析框架, 细化分析应用的关键技术, 提出大数据应用软件实现的要点, 为水质监测工作带来帮助。

**关键词:** 数据存储 数据处理 水利信息建设 自动监测功能

**中图分类号:** TP311.1; X832 **文献标识码:** A **文章编号:** 1003-9082(2023)02-0001-03

## 引言

近几年的发展中, 国家建设逐渐加大了投入力度, 数据信息技术得到了发展, 水利行业受到信息化的影响, 开始形成信息化综合体系, 将大数据软件应用在水利中, 促进了水利工程管理的效率, 利用新型的数据处理模式, 改善着水利信息建设的水平, 推动了水利行业进步, 保证水利基础设施建设更加完善, 逐渐朝着智能化的方向发展。

### 一、水利大数据的特点

信息技术在发展, 遥感设备、传感网等技术形式出现在社会中, 对于水利行业而言, 整体的数据采集水平在提升, 可以获取到更加广泛的数据信息。通过观察这些信息, 可以看出它们的特点繁多, 主要体现出多源异构的形式, 而且分布比较广。结合数据的类型, 包含了大规模的实时检测信息, 主要来自于物联网设备, 比如, 水文气象信息、水位流量变化信息、生态水质等, 也有一些基础信息, 比如, 河流信息、水利工程信息, 这些信息属于档案化形式, 还有一其他行业的有关于水的信息等。在这些数据中可以看出, 并不完全是单独的, 它们之间存在着繁杂的关系, 体现在业务和逻辑方面, 如气候数据发生波动时, 带动水资源量产生变化, 影响空间分布情况, 还会影响水利工程, 不利于水资源的合理分配。观察数据格式, 水利数据规模比较大, 不但包含原来的结构形式的数据, 还会涉及到图品形式、语音形式、视频形式等, 这些属于非结构形式的数据, 致使当前的技术不能良好的处理这些异构数据。从价值密度层面看, 当前物联网设备发展迅速, 遥感技术也获得了广泛应用, 在信息感知方面展现了强大的效果, 衍生了巨大的信息量, 然而, 观察信息的价值密度, 可以看出水平并不高。结合时效性这一特点, 对于某些水利数据, 像洪涝这些紧急发生的情况, 需要第一时间预警, 高效的处理信息, 所以, 需要引进大数据技

术, 保证数据整体处理水平更高。应用水利大数据分析的方式, 可以跨越行业, 也可以跨越部门, 利用多个角度去分析数据, 有效应对随机、大规模的数据, 保证水利工作有序推进。

### 二、基于大数据, 研究其分析框架及关键技术

应用大数据处理平台时, 主要利用的是Hadoop, 然后可以借助MapReduce, 跨越计算机集群设备, 高效的传输数据, 可以整体的提升数据的运算水平。借助水利大数据分析处理方式, 发挥Hadoop的功能, 通过它的分布形式文件系统, 将过往的监测数据存储起来, 还可以存储视频、图片等形式, 对于这些非结构形式的数据, 还有那些半结构形式的数据, 需要合理的执行处理过程, 展现MapReduce的作用, 对于过往监测形成的数据, 执行批量计算和处理, 通过监测得到的数据, 它可以提供分析挖掘作用, 同时也可以利用模型计算<sup>[1]</sup>。在水利多项业务中, 涉及到大量数据, 需要实时处理这些内容, 比如, 雨水情况、水资源情况执行监控形成的数据, 确保这些数据可以准确获取, 才能保证应急事件第一时间得到处理, 从而进行下一步的决策。为了保证数据应用及时, 通过水利数据中心, 完善其架构, 添加Storm计算框架, Hadoop比较适用于处理离线的数据, 而这一数据框架与其不同, 它的数据源可以是更新状态下的, 也就是说, 如果获取到一条数据, 就会及时的处理一条。借助Storm, 可以不限次数处理数据流, 可以随时处理Hadoop的批量任务, 推动专业模型计算工作。

#### 1. 数据采集

进行水利大数据分析时, 首先, 应该针对于来自于不同源头的异构数据, 展开统一集中处理, 可以将采集数据上报处理, 也可以和其他系统有关, 采集其系统的节点数据, 另外包括处于其他领域的交换数据。对采集数据上报, 主要是便于随时监测数据, 接入方式比较灵活, 可以

借助设备,实行直连的方式,也可以利用现场服务器,应用它的转发形式;对于其他系统的节点数据,主要是处于水利业务范围里,具备已经建立完成的应用系统,业务数据分为定时接入以及不定时接入的形式;其他领域的交换数据,主要是其他行业和水利有着一定的联系,从而产生有关的数据,例如,国土数据信息、气象数据信息等,凭借已经定义完善的接口,保证接入到位,或者是定义完善的连接形式,实行接入处理。

### 2. 数据存储

通过水利大数据分析的架构,不仅连接了多层面的数据库数据,还促进了数据库数据和分布形式的文件系统有效接触,为水利大数据的存储带来支撑作用。将大数据技术合理引进,促进了水利数据处理的效果,在监测过程中,应该体现实时的特点,同时应具有自己的结构,增进它和业务处置数据的联系,合理的应用关系型数据库达到存储的效果,对于那些属于半结构形式和非结构形式的数据,例如,以往的监测数据、图像数据等,可以存储在Hadoop分布形式的文件系统中,关于这两种存储方式,它们不是单独存在的,可以借助数据轴达到转换的效果,也可以利用装载工具,实现彼此补充的功效。对于水利大数据,在建设数据库的阶段中,利用BI元数据库形式,有选择的存储数据,主要是那些半结构形式、非结构形式的数据,存储它们的描述和定义内容,应用元数据技术进行定义处理并对元数据进行集中管理,保证系统的实用性更高。

### 3. 数据处理分析

水利大数据分析,主要是依据应用需求,有针对性的处理和分析数据,需要利用并行计算方式,或者是利用云计算体系,凭借其中的数据挖掘技术、机器学习技术等,展开细化分析。在Hadoop中,存储着大量的水利数据,接下来可以利用MapReduce,进行数据分析和计算。MapReduce的功能强大,主要是基于大型的大数据集,实行分解处理,得到多个独立的小数据段,将其分发到多个节点处,利用较短的时间展开并行计算,缓解了单机处理的弊端,提升了计算性能。利用Hadoop的存储功能,再结合MapReduce的计算能力,形成了一套数据库的处理机制,对于大量的半结构形式数据,通过映射处理为表,自动的进行数据处理。凭借Hadoop的Hive给予的接口,可以减少作业中的难度。在水利数据中,还存在着实时监测数据,它需要实时计算处理并显示出来,通过Storm,对数据库进行处理和更新,利用较短的时间计算大规模的监测数据,体现着实时性的作用,便于预警处理、及时展示监测的数据。

### 4. 数据应用

水利数据规模比较大,涉及到的范围比较广,可以结合需要构建通用性的服务内容,也可以建立定制形式的服务,例如,利用水利工程安全监测体系、评价系统、维护工具等,发挥这些安全管理应用内容的作用,沿着水资源多目标进行优化处理,如,预防洪水方面、发电方面、航运方面、农业方面等。通过分析用户数据,深度挖掘业务有关数据,凭借智能化的特点,获取用户兴趣比较浓的数据指标。

### 5. 与原始的水利数据分析方法对比

传统的水力数据分析技术较为落后,与当前的社会发展不匹配。例如,在某水利工程企业中,由于其发展时间较长,企业的跨度较大。因此,其见证了近几十年来水利工程在水利数据分析方法上的革新。在过去,该企业使用传统的调查取样法,该方法在进行水利数据分析时,不仅效率慢,人工量大,而且所调查的数据经常出现不准的情况。其采用抽样调查的过程与组织形式较为传统,一般情况下,该企业开展抽样调查是以部门组织的形式开展的,对于部门人员数量有较大的要求,因此,需要一定的工作人员来补充,协助部门展开数据调查。在工作效率上还有待提高,在水利数据的抽样调查中,由于工作量较大,部门成员在及工作中难免出现工作效率低下的状况。同时,在技术手段上迟迟得不到创新,对于部门成员的工作积极性也难以有效的提高。另外,在传统的水利数据调查中,进行数据分析的角度较为单一,对于事件的进展难以进行有效的预测,这都受限于当时的技术条件,与水利工程的发展思维<sup>[2]</sup>。

而当下在进行水利工程数据分析时,该企业早在大数据刚刚兴起的时期,对未来企业发展作出判断。通过对数据分析与处理相关技术与手段的开发应用,逐渐在水利工程的发展上拔得头筹。首先,在技术手段上,该企业为实现节省大量的人力资源,主动进行自动化、信息化的建设。将一些能用智能机器替代人工操作的环节进行优化,有效的解决人工操作所带来的不确定性。例如,在数据监测环节,使用大数据进行检测能有效避免人工出错。进而在水利工程数据分析中,部门的组织形式得到了持续的优化,对一些需要跨部门,跨企业之间的数据交流节省了工作时间,提高了工作的效率。同时,水利大数据平台建设,对于多角度分析数据与未来走势有着非常深远的意义。避免了过去因部门之间很难进行及时有效的沟通,导致出现数据处理出错的局面。另外,在大数据时代,计算机程序的

计算相较于传统的数据计算成本较低,而且随着信息时代不断发展,越来越多的先进技术得到创新,计算能力与计算软件也在不断发展。

### 三、大数据分析的软件实现

#### 1. 水质自动监测信息

大数据时代下,水利工程水质的检测是保护水资源安全的重要手段。传统的水质检测工作较为复杂,对于人工需求量较大,同时技术手段相较匮乏。导致在水质监测中对信息收集不全,很难真实客观的反映出真实的水质状况。另外,在传统的水质安全监测中,对于自动化技术的使用较少,因此,效率较低。例如,在我国某大型水质监测站中,通过近几年大数据技术在水利工程中的应用,水质监测与信息的收集上有一定的优势。同时,该水质监测站的工作人员也较少,使用的设备较为先进。一般情况下,只有较少的几名工作人员轮岗,而监测机器则是通过水质监测机器进行全天的水质状况信息收集。在收集水质信息中,主要包括以下几点,第一,水质的酸碱度,溶解氧的行亮、高锰酸钾指数等。同时,对于水质中的化学元素含量进行监测<sup>[1]</sup>。

#### 2. MapReduce 计算模型分析

为充分了解水质情况,某水质监测站通过建立 MapReduce 计算模型来实现对水质数据信息的收集与分析。这是一套较为完善的水质监测大数据平台,为水质监测站的工作创造了较为便利的条件。同时,该计算模型在进行数据分析与储存上有一定的优势。MapReduce 计算模型的运行主要包括两个阶段,第一,对水质中历年来的数据信息进行统计,并记录保存。在数据统计阶段,该计算模型自动将数据进行分组,并将内容分组储存。第二,在水质信息数据进行分组记录与保存后,进行下一步操作,将详细信息进行聚合操作,并根据水质监测站的需要进行相关的内容输出。在 MapReduce 计算模型平台中,对于数据的处理过程较为复杂,首先将记录的数据进行函数处理。在该阶段,一般是由计算机系统内部进行自动处理。其次,在数据处理的中间阶段,会针对任务的数量进行相应的减少,进而实现在 MapReduce 计算模型平台中的数据分片,提升了数据的应用率,结合数据结构的存储要求,为多结构数据带来解决路径。

#### 3. 数据准备

在水质监测站数据准备阶段,首先,对水质数据相关内容进行详细的分析,在处理输出后对其进行文本上的保存。保存的原则则是每个行为保存一条的形式,在该数据库中主要储存了 25 条相关的水质类型,对于水质类型的丰

富起到至关重要的作用。同时,在进行数据处理时,还是运用较为规范式的处理方法,这对水质监测站数据处理的真实性就有极强的现实意义。随后水质监测站针对数据文件进行分片处理,在分片处理中,对原则标准的要求,会根据不同的水质类型信息进行适当的调整。在分片数据的参数上,也有较为明显的不同,但都是要将其进行规范化处理,使用函数关系将其进行转化,并使用 Map 数据局技术将其输出。最终,将分片数据 Map 中间结果的数值进行全部的输出<sup>[4]</sup>。

#### 4. 结果分析

在水质分析数据来源中,需要将近年来当地所有的水质监测站的数据进行收集。并进行结果分析,该水质监测站将近几年的数据进行收集,并与分析得出的数据结果进行对比研究<sup>[5]</sup>。其会采用一吨的水量进行水质状况的分析,而数据对比的时间一般会用近三年的数据进行对比。另外,在进行大数据平台建设时,相关领域会主动求变,在分析程序的运行时,将数据储存作为其发展的目的。该水质监测站在数据的处理与储存上进行管理方式的创新,使得水利大数据在应用上,逐渐与其它技术进行融合,例如,智能机器算法领域,数据融合技术等。为后续在水利大数据相关软件的开发应用上实现突破。

### 结语

在大数据技术的支持下,促进水利数据得到高效采集,方便展开数据管理工作,推动水利行业迅速发展。在水利数据的分析处理过程加入大数据技术,可以精准的获取到所需的数据信息,提升数据的应用率,达到数据结构的存储要求,为多结构数据提供解决路径,突破了原来的搁置问题,利用新型的方式,提升了数据处理的效率。

### 参考文献

- [1]全占东.基于水文水资源数字化监测模式研究[J].黑龙江水利科技,2022,50(07):205-207.
- [2]梁志锋,王信堂,刘鑫鹏.基于大数据架构的水文监测平台关键技术研究[J].信息技术,2022(06):6-11.
- [3]牛智星,嵇海祥,崔彦萍.水文监测数据统一接收软件设计与应用[J].水利技术监督,2022(02):43-45+122.
- [4]谢宇扬.云平台下的水文监测数据共享模式研究[D].西安:西安科技大学,2019.
- [5]牛睿平,刘筠.大数据技术在水文监测中的软件应用研究[C]//中国水利学会2019学术年会论文集第二分册.2019:470-475.

作者简介:王爽(1993.03—),女,汉族,河南南阳人,职称:助理工程师,学历:本科,研究方向:水文。